NUCLEIC SEQUENCE AND DEDUCED PROTEIN SEQUENCE FAMILY WITH HUMAN ENDOGENOUS RETROVIRAL MOTIFS, AND THEIR USES

The present invention relates to a novel nucleic sequence and deduced protein sequence family with complete or partial human endogenous retroviral motifs, and sequences flanking or adjacent to said sequences, and controlled by the latter; modification of the expression or impairment of the structure (polyadenylation, alternative splicing and the like) of said flanking sequences.

The invention also relates to the detection and/or use of said nucleic sequences and of said corresponding protein sequences in the context of diagnostic, prophylactic and therapeutic applications, in particular for neuropathological conditions with an autoimmune component such as multiple sclerosis.

The invention also relates to the production of antisense double-stranded and single-stranded nucleic probes, of ribozymes, capable of modulating viral replication (T.R. Cech, Science, 1987, **236**, 1532-1539; R.H. Symons, *Trends Biochem. Sci.*, 1989, **14**, 445-450) of the corresponding recombinant molecules, and associated antibodies.

Retroviruses are viruses which replicate solely by using the opposite route to the conventional processing of genetic information. This process, called reverse transcription, is mediated by an RNA dependent DNA polymerase or reverse transcriptase, encoded by the *pol* gene. Retroviruses also encode at least two additional genes. The *gag* gene encodes the proteins of the skeleton, matrix, nucleocapsid and capsid. The *env* gene encodes the envelope glycoproteins. Retroviral transcription is regulated by promoter regions or "*enhancers*" situated in highly repeated regions or LTR (*Long Terminal Repeat*) and which are present at both ends of the retroviral genome.

During the infection of a cell, polymerase makes a DNA copy of the RNA genome; this copy may then

integrate into the human genome. Retroviruses do not kill the cells which they infect, but on the contrary often enhance their rate of growth. Retroviruses can infect germ cells or embryos at an early stage; they can, under these conditions, integrate the germ line and be transmitted by vertical Mendelian transmission, which constitutes the closest relationship between a host and its parasite. These endogenous viruses can degenerate during generations of the host organism and lose their initial properties. However, some of them may conserve all or part of their properties or of the properties of their constituent motifs, or acquire novel functional properties having an advantage for the host organism, which would explain the preservation of their sequence.

The existence of endogenous motifs having long open reading frames and/or subjected to a strong selection pressure can therefore be an indication of a preserved or acquired biological function, which may correspond to a benefit for the host organism. These retroviral sequences can also undergo, over the generations, discrete modifications which will be able to trigger some of their potentials and generate or promote pathological processes. It has recently appeared necessary to carry out a review and to identify these sequences so as to be able to evaluate their functional impact.

Human endogenous retroviral sequences or HERVs represent a substantial part of the human genome. These retroviral regions exist in several forms:

- complete endogenous retroviral structures combining *gag*, *pol* and *env* motifs, flanked by repeat nucleic sequences which exhibit a significant analogy with the LTR-*gag*-*pol*-*env*-LTR structure of infectious retroviruses,

- truncated retroviral sequences; for example the retrotransposons lack their *env* domain and the retroposons do not possess the *env* and LTR regions.

Up until now, the study of these regions of the genome has been neglected in humans for essentially two reasons:

- the existence of insertions/deletions which can shift the reading frame and of mutations which modify the sequence. These modifications cause impairment of the structure and consequently of the biological function of these motifs,

- the absence of confirmed associations with human pathological conditions.

The recent knowledge of fragments which are significantly representative of the human genome and an orientation of research studies toward a study of structure/function of endogenous retroviral motifs have made it possible to specify the importance of these regions. The involvement of truncated or complete endogenous sequences in pathological conditions in animals is documented; for example their association with tumor processes has been clearly demonstrated (S.K. Chattopadhyay et al., 1982, Nature, **295**, 25-31). Research aimed at specifying the association or the influence of HERVs in human pathological conditions is now therefore justified.

A classification of the HERV elements has been proposed (Tönjes R.R. et al., *AIDS & Hum. Retroviral.*, 1996, **13**, p261-p267; A.M. Krieg et al., FASEB J., 1992, **6**, 2537-2544). It is based on a homology of these sequences with retroviruses isolated in animals, with the aid of heterologous retroviral probes. Indeed, in general, the HERVs exhibit relatively little homology with known human infectious retroviruses.

The class I families exhibit a sequence homology with the type C mammalian retroviruses; there may be mentioned in particular the ERI superfamily, close to the MuLV virus (*murine leukemia virus*) and to the BaEV virus (*baboon endogenous virus*).

The class II families exhibit a sequence homology with the type B mammalian retroviruses such as

MMTV (*mouse mammary tumor virus*) or the type D retroviruses such as SRV (*squirrel monkey retrovirus*).

Other families have also been described; among these, there may be mentioned HERVs which exceptionally exhibit partial homology with HTLV-1 (RTVL-H) or primate viruses; HRES-1, for example, exhibits sequence homology with HTLVs.

Programmes for very large sequencing of the human genome now make it possible to have available a significant number of novel retroviral sequences. The use of data processing software packages makes it possible to identify and analyse these genes. In this context, a systematic search relating to the entire information available to date has been initiated in order to identify novel human endogenous retroviral sequences as a function of certain analytical criteria:

- presence of long open reading frames conserved during evolution of the host organism and which may suggest a biological function,
- analogy with sequences already characterized outside or inside the retrovirus domain,
- location in regions of susceptibility for certain pathological conditions or close to essential genes, for example in the cancer domain, regulation of the immune system or in certain neuropathological conditions.

The work carried out by the inventors on sequence databases allowed them to identify a set of endogenous retroviral sequences or motifs whose normal or pathological expression can promote or disrupt a protective effect in relation to pathological processes, or play a role in the onset or worsening of pathological conditions.

The subject of the present invention is a purified nucleic acid fragment, characterized in that it comprises all or part of a sequence encoding a human endogenous retroviral sequence, which has at least *env*-type retroviral motifs, corresponding to the sequence SEQ ID NO: 1 or to a sequence exhibiting a level of

homology with said sequence SEQ ID NO: 1 greater than or equal to 80% on more than 190 nucleotides or greater than or equal to 70% on more than 600 nucleotides for the *env*-type domains.

The expression homologous sequence is understood to mean both a sequence which exhibits complete or partial identity with the abovementioned sequence SEQ ID NO: 1 and a sequence which exhibits partial similarity with said sequence SEQ ID NO: 1.

According to an advantageous embodiment of said fragment, it has retroviral motifs corresponding to an *env* domain and corresponding to the sequence SEQ ID NO: 1 and retroviral motifs corresponding to a *gag* domain and corresponding to the sequence SEQ ID NO: 2 or to a sequence exhibiting a level of homology greater than or equal to 80% on more than 190 nucleotides or greater than or equal to 70% on more than 600 nucleotides for the *env*-type domains and a level of homology greater than or equal to 90% on more than 700 nucleotides or greater than or equal to 70% on more than 1 200 nucleotides for the *gag*-type domains, said motifs having no insertion or deletion of more than 200 nucleotides.

Said fragments constitute a novel family of human endogenous retroviral sequences (HERV-7q family) which exhibits sequence homology with the MSRV retroviruses, as described in International Application WO 97/06260; said fragments according to the present invention have:

- two repeat nucleotide motifs of 711 bp (Figure 3), having characteristic signals identified in LTRs (*Long Terminal Repeats*): transcription promoters of the TATAA or CCAAT box type. These repeat domains delimit three deduced motifs of the *gag, pol* and *env* type (Figure 2),

- an *env*-type motif (positions 6965 nt - 9550 nt on the sequence SEQ ID NO: 3 or in Figure 1) which contains a long open reading frame of 1 620 nucleotides (positions 7874-9493 of the sequence

ID NO: 3 and Figure 1) encoding a protein having an unpublished sequence of 540 amino acids called enverin (Figure 4 and SEQ ID NO: 26) and underlined fragment in Figure 18. There is present inside the transmembrane domain of this *env* domain a peptide motif of the CKS-25/CKS-17 type (Figure 5), recognized as having immunosuppressive functions on the host lymphocytic cells (M. Mitani et al., 1987, *Proc. Natl. Acad. Sci.* USA, **84**, 237-240). A *zinc finger* type domain **$HX_{3-4}HX_{22-33}CX_2C$** (Kulkolski et al., 1992, *Mol. Cell. Biol.*, **12**, 2331-2338), which is present in integrase-type domains is identified in another reading frame. This particular *env* domain signatures the characteristic of novel endogenous retroviral motifs,

the motif (positions 3065 nt - 4390 nt on the sequence SEQ ID NO: 3) of the *gag* type encoding protein motifs according to Figure 6 (SEQ ID NO: 58) (positions 3118-4198 of SEQ ID NO: 3) was identified by virtue of analogies with known *gag* domains. The region of major homology **QX_3EX_7R** is for example present (Benit et al., 1997, *J. Virol.*, **71**, 5652-5657). The nucleic acid binding motif **$CX_2CX_{3-4}HX_4C$**, situated at the C-terminal position, is identified in another reading frame (Covey et al., 1986, *Nucleic Acids Res.*, **14**, 623-633). Upstream of the *gag* domain, a motif of 182 nucleotides is detected which is repeated twice (Figure 1),

- the *pol* domain exhibits the conventional consensus of a retrovirus *pol* region at the level of the protease, reverse transcriptase and RNAse H domains. A motif close to the consensus **LLDTGA** is found in *pol* (Weber et al., 1988, Science, **243**, 928-931). The motifs **D** and **AF**, **LPQ** and **SP**, and **YVDD** (Xiong and Eickbush, 1990, EMBO J., **9**, 3353-3362) are respectively found in the 3rd, 4th and 5th homology boxes. The motifs **YTDGSS** and **TDS** are present in the RNAse H region,

- the *gag* and *pol* regions could be considered as being joined with a passage from the *gag* region to the *pol* region by a reading frame shift.

The present invention includes the sequences belonging to the HERV-7q family as defined above (presence of the SEQ ID NO: 1 sequence or of a homologous sequence or presence of both the sequences SEQ ID NO: 1 and SEQ ID NO: 2) and in particular the sequences SEQ ID NO: 3-22, 28 and 61; it also includes the complementary nucleic sequences and the reverse sequences complementary to the preceding sequences as well as fragments derived from the coding regions of the preceding sequences corresponding to a shifting frame greater than or equal to 14 nucleotides or their complementary sequences (SEQ ID NO: 37-57, 59-60 and 121-122).

These various fragments may be advantageously used as primers or as probes (reagents A); they hybridize specifically under high stringency conditions to a sequence of the HERV-7q family.

Among these fragments, the following fragments may be preferably mentioned:

- a fragment of 182 nucleotides, repeated twice, situated upstream of the *gag* domain at positions 2502-2611/2613-2865 of SEQ ID NO: 3:

<u>Primers and probes specific for the *gag* region</u>

- a sense primer G1F located in the region upstream of the *gag* domain of *HERV*-7q: 5'GGACCATAGAGGACACTCCAGGACTA3' (SEQ ID NO: 37);

- an antisense primer G1R located in the terminal 3' region of the *gag* domain: 5'CCTCAGTCCTGCTGCTGGATCATCT3' (SEQ ID NO: 38)

- the fragment of 1505 nt amplified by the pair G1F-G1R is used in order to generate the probes capable of hybridizing the various PCR amplification products:

  - a nested sense primer G2F: (SEQ ID NO: 39) 5'CCTCCAAGCAGTGGGAGGAAGAGAATT3'
  - a nested antisense primer G2R: (SEQ ID NO: 40) 5'CCTTCCCTGTGTTATTGTGGACATCATT3'
  - a nested sense primer G4F: (SEQ ID NO: 41) 5'GGAAGAAGTCTATGAATTATTCAATGATGT3'
  - a nested sense primer G3F: (SEQ ID NO: 42)

5'GGGACACAGAATCAGAACATGGAGATT3'

- a nested antisense primer G4R: (SEQ ID NO: 43)

5'GCCTTCAGAAGAGTCAGGTGACAGAGA3'

- a nested antisense primer G5R: (SEQ ID NO: 44)

5'GAGCCTCCAAAGTCCACTTGCCTGA3'

Primers and probes specific for the *env* region

- a sens primer E1F: (SEQ ID NO: 45)

5'GATTTCAGTATCTACTAGTCTGGGTAGAT3'

- an antisense primer E1R: (SEQ ID NO: 46)

5'CTAGGAAATCCAGCTAGTCCTGTCTCA3'

- the fragment of 2529 nt, amplified by the pair of primers E1F-E1R, is used to generate the probes capable of hybridizing the various PCR amplification products:

- a sense primer E2F: (SEQ ID NO: 47)

5'CCAAGACAGCCAACTTAGTTGCAGACAT3'

- an antisense primer E2R: (SEQ ID NO: 48)

5'GGACGCTGCATTCTCCATAGAAACTCTT3'

- a sense primer E3F: (SEQ ID NO: 49)

5'GCAATACTACATACACAACCAACTCCCAA3'

- an antisense primer E3R: (SEQ ID NO: 50)

5'GGGGGAGGCATATCCAACAGTTAGTA3'

- a sense primer E4F: (SEQ ID NO: 51)

5'CCATCTACACTGAACAAGATTTATACACTT3'

- an antisense primer E4R: (SEQ ID NO: 52)

5'AATGCCAGTACCTAGTGCACCTAGCACT3'

- a sense primer E5F: (SEQ ID NO: 53)

5'CGAATACAACGTAGAGCAGAGGAGCTTCGAA3'

- a sense primer E6F: (SEQ ID NO: 54)

5'AGCCCAAGATGCAGTCCAAGACTAAGAT3'

- a primer E5R: (SEQ ID NO: 55)

5'GCGTAGTAGAGGTTGTGCAGCTGAGAT3'

- a primer ExF: (SEQ ID NO: 56)

CCCTTACCAAGAGTTTCTATGGAGAAT

- a primer ExR: (SEQ ID NO: 57)

ACCGCTCTAACTGCTTCCTGCTGAATT

All the oligonucleotides are designed to be able to generate a sense primer and an antisense primer by a shift in the sequence of the reference primer of 1 to 7

nucleotides toward the 5' side or toward the 3' side; the modification of the sequence may cause a modification of the size of the primer of 1 to 7 nucleotides depending on the cases. The primers chosen may be optimized depending on the cases by shortening or extension affecting 1 to 9 nucleotides.

Preferably, the hybridization, cloning, subcloning, production, preparation and analysis of the nucleic acids, peptides and antibodies, the sequencing of the nucleic acids and peptides, the *in situ* hybridization and the immunohistochemistry are carried out under the conditions described in the following books:

- Current Protocols in Molecular Biology, Eds. F.M. Ausubel, R. Brent & R.E. Kingston et al. Green Publishing associates and Wiley Interscience.

- Molecular Cloning: a laboratory manual. Eds. J. Sambrook, E.F. Fritsch & T. Maniatis, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

- The Practical Approach series. Eds. D. Rickwood & B.D. Ames, IRL Press and Oxford University Press. In particular antibodies I & II; DNA cloning I, II, III; Nucleic acid and protein sequence analysis; Nucleic acid hybridization; Nucleic acid sequencing; Oligonucleotide synthesis; Protein purification applications; Protein purification methods; Protein sequencing; Transcription and translation; Gels electrophoresis of nucleic acids; Gels electrophoresis of proteins; Genome analysis; HPLC of macromolecules; Human genetic diseases; Microcomputing in biology; Molecular neurobiology; Mutagenicity testing; Essential molecular biology I & II.

- Proteome research: New frontiers in functional genomics, Eds. M.R. Wilkins et al., Springer.

The human endogenous retroviral sequence (SEQ ID NO: 3) situated on the long arm of chromosome 7 corresponds to the HERV-7q sequence; it has 10.5 kb

(Figs. 1 and 2) and satisfies the criteria defined above.

The search for domains exhibiting total or partial similarity with the *gag* and *env* regions of HERV-7q resulted in the identification of novel endogenous retroviral sequences. These sequences may have the structure of a complete endogenous retrovirus such as the endogenous retroviral sequence situated close to the gene for the alpha and delta subunits of the T cell receptor, and consequently called HERV-TcR; by way of example, Figure 7 shows the comparison of the nucleic alignments of the respective *gag* domains of HERV-7q and HERV-TcR (sequence HG12, SEQ ID NO: 19). Partial retroviral structures also exist. These retroviral domains, similar to HERV-7q, are identified in independent nucleic sequences as shown by their chromosomal location. Nucleic motifs (called here HEx or HGx, and analogous to *env* or *gag* type domains, respectively) resembling the *env* or *gag* domains of HERV-7q were found, with the aid of the abovementioned databases:

- HE2: chromosome 17 (SEQ ID NO: 4),
- HE3 and HG3: chromosome 6 (SEQ ID NO: 5 and 6),
- HE4: chromosome X (SEQ ID NO: 7),
- HE5: chromosome X q22 (SEQ ID NO: 8),
- HE6 and HG6: chromosome 1 q23.3-q24.3 (SEQ ID NO: 9 and 10),
- HE7: chromosome 7 p15 (SEQ ID NO: 11),
- HE8 and HG8: chromosome 19 (SEQ ID NO: 12 and 13),
- HE9: chromosome X (SEQ ID NO: 14),
- HE10: chromosome X q13.1-21.1 (SEQ ID NO: 15),
- HE11 and HG11: chromosome 7 q21-22 (SEQ ID NO: 16 and 17),
- HE12 and HG12, in HERV-TcR: chromosome 14 q11.2 (SEQ ID NO: 18 and 19),
- HE13 (SEQ ID NO: 61): chromosome 6 q24.1-24.3

The present invention also includes the coding and noncoding fragments for all or part of enverin

comprising at least 14 nucleotides and in particular the fragments encoding the C-terminal part of enverin, either from amino acid 291, or from amino acid 321, starting from the first methionine.

These fragments comprise in particular a critical zone where two inserts of 12 nucleotides were characterized:

- a first insert was identified (sequence A) in individuals of 2 groups (patients and controls). This insert, situated between amino acids 487 and 488, makes it possible to insert the tetrapeptide VLQM. A comparative analysis shows that this insert is identified in a homologous region situated in the sequence HE13, belonging to the HERV-7q family. The amplification of the HE13 type sequence could indicate that there is an impairment of the enverin sequence of HERV-7q, which would promote the amplification of the sequence contained in HE13. This observation also makes it possible to use this insert as a specific element for amplification of sequences of the HE13 type.

A second insert (sequence B) was identified in a patient with MS. The insert of 12 nucleotides is situated at the level of amino acid 495 and encodes the tetrapeptide MQSM. It is remarkable to observe that this insert is also identified in a homologous region situated in HE13.

Sequence A: TAAACTACAAATGGTTCTTCAAATGGAGCCCA (SEQ ID NO: 59)

Sequence B: GATGCAGTCCAAGATGCAGTCCATGACTAAGA (SEQ ID NO: 60).

These observations demonstrate modifications of the enverin sequence of the HERV-7q type which constitute the basis for a detection strategy by allele-specific amplification (AS-PCR), making it possible to detect these differences in a population and which could correspond either to a mutation/deletion associated with a degree of susceptibility, or to a polymorphism, or to a

mutation/deletion associated with a pathological condition such as multiple sclerosis.

The alignments of the *env* (Fig. 8) and *gag* (Fig. 9) domains explain the levels of homology observed between the sequences described above and the homologous sequences in HERV-7q. The analogies can extend to the flanking retroviral motifs.

Analysis of the sequence tags available in databases shows that transcripts belonging to some members of this family, in particular HERV-7q, are essentially expressed in tissues of foetal or placental origin.

Polypeptide sequences generated by these transcripts can therefore be potentially produced and biological functions or activities can be envisaged, by analogy with biologically active polypeptides of viral or retroviral origin; for example, the peptide motifs of the CKS-17 type (Haraguchi et al., PNAS, 1995, 92, 5568-5571) (Fig. 5) or CKS-25 type (Huang S.S. and Huang J.S., J. Biol. Chem. 1998, 273, 4815-4818) which have immuno-modulatory functions on the lymphocytic host cells. The differences in sequence which are observed and possible normal or pathological modifications are in particular responsible for modulation of the function.

HERV-7q represents the paradigm of the novel family of human endogenous retroviral sequences or of endogenous retroviral motifs.

HERV-7q and some of the endogenous retroviral sequences belonging to its family have a *pol*-type domain analogous to *pol*-type retroviral sequences such as for example the *pol* region identified in the MSRV retrovirus associated with multiple sclerosis and described by H. Perron et al. (1997, *Proc. Natl. Acad. Sci.* USA, **94**, 7583-7588; International Application PCT WO 97/06260).

However, the sequences according to the present invention are distinguishable from the infectious exogenous retroviral sequences analogous to MSRV

previously described in that the *gag* and *env* sequences according to the invention are significantly different according to the criteria defined above and as a function of certain specific characteristics, for example the long open reading frame of the *env* domain of HERV-7q; they would be able to allow the signaturing of a pathological condition when they have insertions, deletions, reading frame shifts or mutations.

Indeed, the differences observed between the human sequences of the HERV-7q type, which are isolated from individuals reputed to be normal, and the sequences derived from some samples of pathological origin are not randomly distributed. Comparisons carried out between the *gag* region obtained from infectious retroviral particles (EMBL accession No.: A60168, A60200, A60201, A60171 and the like) and the corresponding *gag* sequence of HERV-7q (Fig. 9), make it possible to observe that the mutations preferably affect non-sense codons. For example, two non-sense codons in HERV-7q are replaced by an arginine codon in A60200, which makes it possible to obtain a deduced sequence of 109 amino acids for HERV-7q and of 166 amino acids for A60200. The base changes consequently make it possible to extend the reading frame and to potentially encode larger sized polypeptide structures (Figure 10).

Likewise, an *env*-type sequence obtained from infectious retroviral particles exhibits a significant analogy with the *env* domain of HERV-7q (Figure 11). These marked analogies between exogenous and endogenous retroviral sequences could be responsible for the triggering or worsening of certain pathological processes, in particular certain autoimmune diseases such as multiple sclerosis. In this regard, it is possible to note that certain endogenous retroviral sequences described in the invention are situated close to or in regions reputed to exhibit susceptibility for multiple sclerosis: for example HERV-7q and the 7q21-22 region of chromosome 7, likewise for HE12 and HG12 in

HERV-TcR and the region of the gene encoding the alpha and delta chains of the T cell receptor, HE2 and chromosome 17, or HE3, HE13 and HG3 and chromosome 6, for example, the sequences HE11 and HG11, around the region 7q 21-22 or HE4, HE5, HE6, HE9, HE10 or HG10 on the X chromosome. These sequences would therefore be capable of providing the means for locating or identifying the genes for predisposition.

No significant homology is observed with endogenous retroviral sequences already described; on the other hand, a limited homology may be noted, which makes it possible to identify a general structure of the *env* domain; however, said homology is less than the criteria defined according to the invention between the *env* domains of the sequence HERV-7q (SEQ ID NO: 1) and the sequence HERV-9 (Figure 12). Figure 11 shows extensive homologies between the sequence HERV-7q with an exogenous retroviral sequence (accession No. EMBL: A60170).

The human endogenous retroviral sequences belonging to the HERV-7q family can protect against attacks linked to the environment or can be beneficial for the individual. This beneficial effect could be one of the possible reasons for the selection pressure exerted on some of these sequences and the potentially functional character of the deduced protein structures identified: for example the long open reading frame capable of encoding a novel protein and corresponding to the *env* domain of HERV-7q.

The human endogenous retroviral sequences belonging to the HERV-7q family could be associated, for example, with pathological conditions related to processes linked to cancer, to neuropathological conditions with an autoimmune component or to any other pathological process in association or otherwise with endogenous or exogenous viruses or retroviruses. Their action could be related to the outbreak, the worsening, the modification of the time of appearance or the protection against the disease.

In the context of application to autoimmune pathological conditions (such as for example lupus, Sjögren's syndrome, rheumatoid arthritis, multiple sclerosis and the like), significant analogies may be detected between the endogenous retroviral motifs identified and motifs found in retroviral structures characterized in patients with autoimmune pathological conditions such as multiple sclerosis; for example, fragments of *gag* domain (recently available in databases) obtained from infectious retroviral particles or the complete sequence of the *pol* domain corresponding to the MSRV virus associated with multiple sclerosis. These retroviral motifs possess significant analogies with homologous endogenous sequences of the HERV-7q type, which makes it possible to envisage direct or indirect association with pathological processes, including multiple sclerosis, in association or otherwise with MSRV.

The importance of these sequences goes beyond the context of autoimmune diseases. Apart from the general importance of retroviral motifs in the triggering or worsening of a tumor process, which is well established in particular in murine models (H. Fan in *The retroviridiae*, 1994, ed. J.A. Levy, Plenum, New York, p. 313-353), these sequences could be present close to or inside important genes and could alter the expression thereof: for example HERV-TcR and the genes for the alpha and delta subunits of the receptor for the T cells involved in disruptions of the immune system.

The present invention includes, in addition, the use of sequences combined with the sequences of the HERV-7q family for the detection and/or prognosis of various autoimmune diseases (neuropathological conditions in particular); these sequences encode all or part of a factor whose function, the regulation/de-regulation or alteration (polyadenylation, alternative splicing), is associated with the normal or pathological expression or with the regulation/de-

regulation of the motifs belonging to the HERV-7q family and correspond to transcripts or cDNAs of the nucleotide sequences encoding genes situated in regions flanking or delimiting retroviral sequences of the HERV-7q family.

The expression flanking region is understood to mean any region situated close to (contained in or including) an endogenous retroviral sequence belonging to the HERV-7q family, as defined above, up to and including the genes immediately contiguous and/or situated at a distance which cannot exceed 120 kb.

The inventors have now found that the presence of the retroviral sequences as defined above disrupts the expression or impairs the structure of the flanking sequences defined below.

The transcripts of said flanking sequences (and fragments thereof, in particular those underlined or in italics in Figures 14-16, 22-26, as defined below:

- at 1021 bp upstream of HERV-7q, there is identified an endogenous retroviral sequence called RH7 (SEQ ID NO: 62 and Figure 22); this sequence is situated in 5′ of the HERV-7q sequence; in Figure 22, the portion in italics corresponds to the beginning of the HERV-7q sequence; the RH7 sequence is underlined; two putative polyadenylation sites are in bold. This sequence SEQ ID NO: 62 exhibits significant homology, on more than 6 kb, with RGH-type endogenous retroviral sequences (Figure 13). Sequences belonging to this family are expressed in particular in patients with rheumatoid osteoarthritis (Nakagawa et al., (1997), Arthritis, Rheum., 40, 627-638). The present invention also includes fragments of the sequence SEQ ID NO: 62, comprising between 14 and 50 nucleotides (used as primers), preferably between 14 and 25 nucleotides, or at least 25 nucleotides (used as probe), which fragments have the following characteristics: the 4 nucleotides of the 3′ end are different from the corresponding motifs of the sequence RGH2 (bottom sequence in Figure 13, GenBank accession No.: D110 18),

- at less than 9 kb upstream of HERV-7q, there is identified the sequence RAM75 (SEQ ID NO: 63 and Figure 14) containing the 24 coding exons (which cover close to 41 kb) of the gene for peroxisomal ATPase PEX1. PEX1, in combination with PEX6, is responsible for the import of peroxisomal proteins and for stabilizing the PEX5 receptor. A disruption/alteration affecting PEX1 is responsible for various neuropathological conditions such as Zellweger syndrome, neonatal adrenoleukodystrophy and the infantile form of Refsum's disease (Reuber et al., (1997), Nature Genet., 17, 445-448). It can be recalled that the main function of the peroxisomes is associated with the metabolism of fatty acids, in particular by β-oxidation processes. Impairment of the gene identified in the sequence RAM75, or of its expression, by modification of the function of the regulatory 5' and 3' regions or by modification of the splicings or of the polyadenylation processes, in particular under the influence of neighboring retroviral motifs, would be able to disrupt the expression and the structure of ATPase and consequently to disrupt one of the peroxisomal functions, in particular the metabolism of lipids, in particular myelin lipids, with consequences for certain pathological conditions, including neuropathological conditions such as multiple sclerosis; the underlined portions (Figure 14) correspond to the 24 coding exons.

The present invention also includes the fragments of the sequence SEQ ID NO: 63, included in the abovementioned 24 coding exons and comprising at least 14 nucleotides.

Analysis of the expression profile (transcripts and proteins) of the sequence RAM75 (SEQ ID NO: 63) is a good indicator for the differential diagnosis of neuropathological conditions with an autoimmune component.

In Figure 14, the coding exons are underlined. The initiation and non-sense codons as well as the

putative polyadenylation sites are in bold and underlined;

- at 0.7 kb downstream of the sequence HERV-7q and on nearly 17 kb (SEQ ID NO: 64 and Figure 15), there is identified the nucleotide sequence RAV73, where there are detected sequence tags and potential exons capable of producing one or more polypeptide sequences; the invention also includes fragments of this sequence SEQ ID NO: 64 included in the sequence tags and the potential exons as they appear (portions underlined) in Figure 15, which fragments comprise at least 14 nucleotides,

- at 120 kb upstream of the sequence HG3, and on 15 kb, there is the nucleotide sequence RBP3 (SEQ ID NO: 65 and Figure 23), which covers the 3' end of the gene encoding a transcription factor of the Blimp-1 family (SEQ ID NO: 119 and Figure 25), a protein of 789 amino acids which is a repressor of the expression of the interferon-beta gene (Keller and Maniatis, Genes Dev., (1991), 5, 868-879), which is already associated with certain malignant pathological conditions (Mock et al., Genomics, (1996), 37, 24-28), and which could play a role in the differentiation and the pathogenesis of B cells. The possible association of the endogenous retroviral sequence containing the motifs HG3 and HE3 and of Blimp-1 has many benefits, in the case of pathological conditions, and in particular multiple sclerosis. Blimp-1 acts in particular on the B cells whose contribution in inflammatory processes associated with multiple sclerosis is known. Blimp-1 is capable of blocking the viral induction of the INFβ promoter whose capacity to reduce the frequency of attacks and the progression of lesions in patients with MS is known. Disruption in the expression or the structure of Blimp-1, in relation to a retroviral element of the HERV-7q type, is consequently associated with neuropathological conditions or with diseases having an autoimmune character, such as multiple sclerosis; this nucleotide sequence RBP3 (SEQ ID

NO: 65) contains nucleotide motifs identified in the nucleic sequence encoding the Blimp-1 gene; the invention also includes the detection of the mRNA sequences for the Blimp-1 protein (SEQ ID NO: 119),

- the endogenous retroviral sequence of the HERV-7q type, containing HE3 and HG3, is situated in the HI3 region corresponding to an intron extending over more than 46 kb (SEQ ID NO: 66), of a gene which could encode the analogue of APS (Figure 24), a protein of 275 amino acids specific to apoptosis, overexpressed in various cells in culture after triggering an apoptotic process (Hammond et al., FEBS Lett., (1998), 425, 391-395). The intron is situated at the level of amino acid 231 of APS. The end of HE3 is at more than 12 kb from the 5' end of the intron, whereas HG3 is situated at more than 28 kb from the 3' end of the intron. Apoptotic processes are associated with multiple sclerosis. In particular, there has been described an apoptotic process affecting astrocytes and oligodendrocytes in the presence of a purified fraction of cerebrospinal fluid of patients suffering from multiple sclerosis (Ménard et al., J. Neurol. Sci., (1998), 154, 209-221).

Finally, it should be stressed that the nucleic region containing HE3, HG3, HI3 and RBP3 is located at the level of the short arm of chromosome 6, in 6p21, which is a proposed region of susceptibility to multiple sclerosis (The Multiple Sclerosis Genetic Group, Nature Genet., (1996), 13, 469-472).

The interaction between the HERV-7q type sequences and the flanking sequences and the importance of establishing a profile of expression including one or more of the abovementioned sequences in order to establish a differential diagnosis of a neuro-pathological condition is even more evident because it is observed that the sequences HG12 and HE12 are situated in an intron region of the gene encoding the alpha and delta subunits of the T cell receptors. The T cell receptors are involved in the immune regulation

process and their influence has been proposed in the case of autoimmune diseases, including multiple sclerosis.

The subject of the invention is also transcripts generated from the abovementioned sequences as well as those optionally exhibiting modifications in the reference sequences described in the invention when they are expressed in certain patients.

Indeed, the systems for regulating the the expression of the retroviral proteins of HERV-7q, which are present in the LTR type motifs, could influence the expression of genes situated in the close or distant chromosomal vicinity and could induce disruptions of an immunological and/or neurological character. For example, the endogenous retroviral sequence HERV-TcR exists in the immediate vicinity of the genes for the alpha and delta subunits of the T cell receptor previously described. The LTR-type motifs could also encode superantigens (Acha-Orbea and Palmer, 1991, *Immunol. Today*, **12**, 356-361). In general, retroviral proteins of the HERV-7q or related type, or their truncated or partial forms could be involved in cytotoxicity or superantigenicity phenomena, such as for example those derived from the long open reading frame identified in the *env* domain (Figure 4).

Sequences of the HERV-7q 5′ and 3′ LTR type, which are highly conserved, are involved in such regulatory effects. By way of example, LTX is described, which is a sequence comparable to that of an HERV-7q LTR (SEQ ID NO: 67 and Figure 16), and which is present in the center of an intron of more than 49 kb, but at 2 kb from the donor 5′ site of the FMR2 gene associated with fragile X and encoding a protein of 1311 amino acids (Figure 26). The LTRs modulate the alternative splicing (Kapitonov and Jurka, (1999), J. Mol. Evol., 48, 248-251), the expression of the gene, the binding to nuclear proteins (Akopov et al., (1998), FEBS Lett., 421, 229-233), or allow the

production of an alternative polyadenylation signal (Goodchild et al., (1992), Gene, 121, 287-294).

In general, there may be noted the existence of several endogenous retroviral sequences of the HERV-7q type (HE4, HE5, HE9, HE10), situated at the level of chromosome X which represents the chromosome associated with the largest number of pathological conditions.

In this regard, it is possible to note that retroviral motifs derived from defective regions are capable of having biological functions; for example, the envelope protein p15E, derived from defective retroviral motifs, possesses an anti-inflammatory and immunosuppressive activity (Snyderman and Ciancolo, 1984, *Immunol. Today*, 5, 240-244).

These structures are probably capable of causing breaks or of amplifying deregulations in the immune defense processes. Some of the motifs of the *gag, env* and LTR-type domains may be associated with a particular function or may contribute to the normal or pathological function of the flanking domains as defined above (SEQ ID NO: 62-67). Recombinations with an element of exogenous, retroviral origin or otherwise can give rise to the production of nucleic or protein motifs which could either protect or trigger or promote or worsen a pathological condition. Likewise, a retroviral structure containing endogenous retroviral elements according to the invention would be capable of causing a pathological process after passing through an exogenous transient cycle followed by reintegration into a sensitive or critical region of the human genome.

It is thus possible to obtain expression profiles (transcripts and optionally proteins) which correspond to the abovementioned neuropathological conditions.

Likewise, the combination of motifs belonging to the HERV-7q family, or of elements induced by motifs belonging to the HERV-7q family, with motifs of exogenous origin or induced exogenously would be

capable of triggering or worsening a pathological process or on the contrary of promoting protection or partial remission or a complete and permanent cure.

The detection made possible of the HERV-7q type domains suggests possible applications at the prophylactic, prognostic and diagnostic level; for example, immunological approaches or gene amplification, which make it possible to compare normal individuals serving as reference with patients, would be capable of promoting screening, of improving early detection of the outbreak of the disease and/or of monitoring the progression of a pathological condition in patients which may exhibit a susceptibility or in whom there has been an outbreak of the disease or in individuals considered to be normal, based on current clinical criteria.

The specific nucleic and immunological probes, as defined, in the present invention are capable of promoting the identification and detection of motifs which are abnormally expressed in the context of pathological conditions associated with cancer, or of neuropathological conditions, in particular autoimmune pathological conditions, at the forefront of which is multiple sclerosis.

The subject of the present invention is also hybrid nucleic sequences, characterized in that they comprise sequences or motifs belonging to the HERV-7q family, or of elements induced by motifs belonging to the HERV-7q family, with motifs of exogenous origin or induced exogenously (exogenous retroviral sequences); such hybrid sequences are probably capable of triggering or worsening a pathological process or on the contrary of promoting protection or partial remission or a complete and permanent cure.

The subject of the present invention is also a diagnostic reagent for the differential detection of complete or partial human endogenous nucleic sequences, having retroviral motifs, selected from the sequences SEQ ID NO: 1 and/or SEQ ID NO: 2, characterized in that

it is selected from the group consisting of the sequences SEQ ID NO: 1-22, 28, 37-57, 59-61 and 121-122, the complementary nucleic sequences and the reverse sequences complementary to the preceding sequences, of nucleotide fragments capable of defining or of identifying the sequences SEQ ID NO: 1 and/or SEQ ID NO: 2 and any flanking sequence or any sequence overlapping them as well as of fragments derived from the coding regions of the sequences SEQ ID NO: 1-22 and 61, corresponding to a shifting frame greater than or equal to 14 nucleotides or their complementary sequences, optionally labeled with an appropriate marker as well as of sequences as defined in Figures 18-21.

The sequences of the nucleic, ribonucleic and oligonucleotide probes used will be chosen from the *env* and *gag* regions or their flanking regions; for example the oligonucleotide primers for HERV-7q will be chosen from the regions situated between nucleotides 3065 and 4390, nucleotides 6965 and 9550 or nucleotides 2502-2865 of SEQ ID NO: 3, as well as from any adjacent sequence (upstream or downstream) capable of allowing specific amplification (Figure 1).

Among the appropriate markers, there may be mentioned radioactive isotopes, enzymes, fluorochromes, chemical markers (biotin), haptens (digoxygenin) and antibodies or appropriate base analogues.

Preferably:

- said reagent is selected from the sequences SEQ ID NO: 37-57 and is capable of being used as a primer,

- said reagent is selected from the following sequences:

a fragment of 1505 nt amplified by the pair of primers SEQ ID NO: 37 and SEQ ID NO: 38 (primers G1F and G1R),

a fragment of 2529 nt amplified by the pair of primers SEQ ID NO: 45 and SEQ ID NO: 46 (primers E1F and E1R),

a fragment of 182 nucleotides, repeated twice, situated upstream of the *gag* domain at positions 2502-2611/2613-2865,

fragments encoding or not encoding all or part of enverin, comprising at least 14 nucleotides and in particular the fragments encoding the C-terminal portion of enverin, either from amino acid 291, or from amino acid 321, starting from the first methionine,

and is capable of being used as a probe.

The subject of the present invention is also a method for the rapid and differential detection of the endogenous retroviral nucleic sequences of the *env* or *env* and *gag* type, their normal or pathological variants, by hybridization and/or gene amplification, carried out using a biological sample, which method is characterized in that it comprises:

(a) a step in which a biological sample to be analysed is brought into contact with at least one probe as defined above, and

(b) a step in which the product(s) resulting from the nucleotide sequence-probe interaction is detected by any appropriate means.

In accordance with said method, it may comprise:

* prior to step (a):

. a step of preparing the relevant biological tissue or fluid,

. a step of extracting the nucleic acid to be detected, and

. at least one gene amplification cycle, and

* subsequent to step (b):

. a step of comparing the nucleic sequences obtained in said biological sample with the human endogenous retroviral sequences according to the invention by any appropriate means and in particular by sequencing, Southern blotting, restriction cleavage, SSCP or any other method which makes it possible to identify an insertion or a deletion or a single mutation between the various sequences compared.

In accordance with the invention, the human endogenous retroviral sequences according to the invention are thus compared with the nucleic sequences present in the biological sample to be analysed and allow the detection of homologous sequences from patients suffering from pathological conditions likely to involve a modification of their genome.

Advantageously, said gene comparisons are carried out using genomic DNA obtained from control individuals and from patients.

A conventional gene amplification by PCR will be carried out with the aid of 5'-sense and 3'-antisense primers delimiting or comprising the zone to be studied (*env* zone or *gag* zone).

Also advantageously, the sequences of the nucleic, ribonucleic and oligonucleotide probes used are chosen from the *env* and *gag* regions or their flanking regions; for example the oligonucleotides which are primers for HERV-7q will be chosen from the regions situated between nucleotides 3065 and 4390 and nucleotides 6965 and 9550, and from any adjacent sequence (upstream or downstream) capable of allowing specific amplification (Figure 1), as specified above. They are preferably selected from the group consisting of

a fragment of 1505 nt amplified by the pair of primers SEQ ID NO: 37 and SEQ ID NO: 38 (primers G1F and G1R),

a fragment of 2529 nt amplified by the pair of primers SEQ ID NO: 45 and SEQ ID NO: 46 (primers E1F and E1R).

The gene amplification step is in particular carried out with the aid of one of the following gene amplification techniques: amplification using Qβ-replicase, PCR, LCR, ERA, CPR or SDA.

The subject of the present invention is also chimeric sequences, characterized in that they consist of a fragment of 17 to 40 nucleotides of a flanking sequence as defined above combined with an endogenous

retroviral motif of the HERV-7q type comprising between 17 and 40 nucleotides, as defined above.

The subject of the present invention is also a method of detecting transcripts as defined above, characterized in that it comprises:

- collecting messenger RNAs obtained from control biological samples (biological tissues, cells or fluids) and from a similar sample collected from patients, and

- the qualitative and/or quantitative analysis of said mRNAs by *in situ* hybridization, by dot-blot, Northern blotting, RNAse mapping or RT-PCR, with the aid of a diagnostic reagent as defined above.

The subject of the present invention is also a method for the detection and/or evaluation of an overexpression/underexpression or of a modification of at least one of the endogenous retroviral sequences or fragments of sequences of the HERV-7q type and/or of their associated flanking sequences, characterized in that it comprises:

- depositing on an appropriate support, such as for example a nylon filter, a glass slide or their equivalent, cDNA or its equivalent obtained from clones, PCR products obtained from genomic DNA, RT-PCR products obtained from transcripts or from specific oligonucleotide sequences, said DNA sequences being endogenous retroviral sequences or fragments of sequences of the HERV-7q type and/or their flanking sequences, as defined above, consisting of transcripts and cDNAs of the genomic sequences, which encode all or part of a factor, whose function, regulation/de-regulation or alteration is associated with the normal or pathological expression or with the regulation/deregulation of motifs belonging to said HERV-7q family, these sequences corresponding to nucleotide sequences encoding genes situated in flanking regions situated upstream and/or downstream of a retroviral sequence of said HERV-7q family and in

which one of the ends cannot be at a distance exceeding 120 kb, and/or a chimeric sequence as defined above,

- the hybridization of said support with at least one appropriately labeled probe obtained, for example, by retrotransposition of an RNA mixture obtained from biological cells, tissues or fluids obtained from controls reputed to be normal, from members of various ethnic populations, from patients suffering from pathological conditions often associated with expression of retroviruses, such as tumor processes, or such as autoimmune diseases, and

- the detection of the hybrids formed.

According to an advantageous embodiment of said method, said transcript or cDNA is selected from the group consisting of the sequences SEQ ID NO: 62-67 and 119 and their fragments corresponding to a shifting frame greater than or equal to 14 nucleotides or their complementary sequences.

According to another advantageous embodiment of said method, said support comprises, in addition, any endogenous or exogenous retroviral sequence.

The method of DNA chips (Bowtell, (1999), Nature Genet., 21, 25-32), is used to evaluate the modification of the expression of all or part of some of the sequences of retroviral origin of the HERV-7q type and flanking sequences. Briefly, DNA obtained from clones, PCR products obtained from genomic DNA, RT-PCR products obtained from transcripts or specific oligonucleotide sequences are deposited on a support, such as for example a nylon filter, a glass slide or their equivalent. The deposited nucleic sequences cover the various retroviral domains described above, as well as the contiguous sequences and the flanking genes. In order to detect possible alternative splicing processes, specific DNAs are synthesized per step of 500-600 nucleotides with an overlap of 250-300 nucleotides on either side. The alternative splicings already identified will be the subject of a specific synthesis. The hybridization is carried out with the

aid of a probe obtained, for example, by retrotransposition of an RNA mixture obtained from biological cells, tissues or fluids obtained from controls reputed to be normal, members of the various ethnic populations, patients suffering from pathological conditions often associated with expression of retroviruses, such as tumor processes, or such as autoimmune diseases, including multiple sclerosis. In this case, a µg fraction and up to a few µg of mRNA or up to a few µg or a few tens of µg of RNA, depending on the method used and the size of the DNA chip involved, are sufficient for the synthesis of the nucleic probe. The nucleic probe is suitably labeled so as to allow subsequent detection, such as for example by fluorescence or by an equivalent method.

The use of bi- or even multicolored probes makes it possible to specify the concerted expression of several genes in parallel, while taking advantage, furthermore, of a precise normalization. The results are acquired automatically, such as for example by a laser scanning system or its equivalent.

Two types of DNA chips are designed, on the one hand chips having an exhaustive set of sequences, and on the other hand specific DNA chips enabling targeting to a more specific application.

For example, a critical sequence in that it would contain a difference relating to a deletion or even a mutation is detected with the aid of specific oligonucleotides (Wang et al., (1998), Science, 280, 1077-1082). The polymorphism associated with a base or with a mutation is detected with the aid of four oligonucleotides possessing one of the four sequence possibilities at the level of a base (A, C, G or T); for each point difference, the 4 oligonucleotides are deposited and the hybridization intensities are compared. Furthermore, an alternative splicing is detected using DNAs corresponding to a single effective or putative exon; the gene is therefore analyzed exon by exon. The DNA chips also relate, by extension, to

any endogenous or exogenous retroviral sequence, such as for example ERV-9, ERV-K, ERV-L, ERV-H, ERV-4, ERV-6, ERV-8, ERV-10, ERV-15, ERV-16, ERV-17, ERV-18, ERV-21, ERV-24, ERV-33, ERV-34, ERV-36, ERV-40, ERV-42, ERV-MLN, ERV-FRD, ERV-FTD and the like), as well as all the putative exon sequences (identified by the existence of sequence tags and corresponding transcripts) or effective exon sequences, and which are situated on either side up to a distance of 120 kb of the endogenous retroviral sequences of the HERV-7q type.

The comparative study is carried out between a control sample and the sample to be tested, in a prophylactic, diagnostic or therapeutic perspective, such as for example the early detection of a modification of the expression of one of the sequences, in a cell, a tissue or an organism, the identification of a sequence associated with a susceptibility or with any pathological condition, the monitoring of the progression of the pathological condition or the monitoring of a treatment and the evaluation of its efficacy.

Apart from the applications already mentioned, the advantage of the method makes it possible, more generally, to make an assessment of the changes observed in an individual, which constitutes to a certain extent an identity card, which facilitates an epidemiological approach which makes it possible to establish novel correlations between a particular observed profile and a pathological condition, in the absence of an *a priori* regarding this pathological condition.

The subject of the present invention is also a kit for the detection and/or evaluation of an auto-immune disease and in particular of neuropathological conditions with an autoimmune etiology, characterized in that it comprises, in addition to the buffers necessary for carrying out the methods as defined above:

- diagnostic reagents A as defined above, and

- reagents B consisting of the transcripts and cDNAs of the genomic sequences, which encode all or part of a factor, whose function, regulation/de-regulation or alteration is associated with the normal or pathological expression or with the regulation/de-regulation of motifs belonging to said HERV-7q family, these sequences corresponding to nucleotide sequences encoding genes situated in flanking regions situated upstream and/or downstream of a retroviral sequence of said HERV-7q family, of which one of the ends cannot be at a distance exceeding 120 kb,

- which reagents are preferably attached to an appropriate support.

According to an advantageous embodiment of said kit, said reagents B are selected from the group consisting of the sequences SEQ ID NO: 62-67 and 119 and their fragments corresponding to a shifting frame greater than or equal to 14 nucleotides or their complementary sequences, as well as the sequences represented in Figures 13-17, 22-26.

The subject of the present invention is also products of translation, characterized in that they are encoded by a nucleotide sequence as defined above.

The subject of the present invention is also a peptide, characterized in that it is capable of being expressed with the aid of a nucleotide sequence selected from the group consisting of the sequences SEQ ID NO: 1-22, 28 and 61, as defined above, according to the combinations offered by the use of the various possible reading frames (see also Figures 18-21).

Said peptide also includes the derived peptides or polypeptides comprising between 5 and 540 amino acids (SEQ ID NO: 23-36 and SEQ ID NO: 58 and their fragments of at least 5 amino acids) and in particular a fragment of 538 amino acids, starting at the first methionine of the sequence SEQ ID NO: 26 (enverin).

According to an advantageous embodiment of said peptides they are in particular selected from the

sequences SEQ ID NO: 23-36, 58, in particular the sequence SEQ ID NO: 26 and its C-terminal fragments, either from the amino acid 291, or from the amino acid 321, starting from the first methionine.

According to another advantageous embodiment of said peptides, they are obtained from nucleic sequences as defined above, in which at least one non-sense codon may be replaced with a codon encoding one of the following amino acids: Phe (F), Leu (L), Ser (S), Tyr (Y), Cys (C), Trp (W), Gln (Q), Arg (R), Lys (K), Glu (E) or Gly (G).

The invention thus includes the deduced peptides or the deduced proteins corresponding to all or part of the nucleic sequences described in the invention, and optionally exhibiting modifications with the reference sequences described in the invention, when they are expressed in some patients. In particular, the invention includes the complete or partial sequences obtained according to the 3 sense reading frames and the 3 reverse and complementary reading frames (see Figures 18-21).

Advantageously, the analysis of the structure of the env domain of HERV-7q, called enverin, made it possible to demonstrate successively:

- an N-terminal signal peptide (region 1-21) and two transmembrane domains (region 320-340; 455-477), responsible for interactions with membrane lipid or protein motifs,

- an immunomodulatory motif of the CKS-17 (Haraguchi et al., (1995), 92, 5568-5571)/CKS-25 type. It is possible to note, in this regard, the presence of an **RalD** motif inside the peptide of the CKS-17/CKS-25 type of HERV-7q and a motif **RvaD** at position 363 which correspond to the consensus W/RxxD, proposed for the active site of the TGF-βs (Huang et al., J. Biol. Chem., 1997, 272, 27155-27159), potent factors associated with growth, with differentiation and with morphogenesis and which are associated with many human pathological conditions, such as tumor processes (Tang

et al., (1998), Nat. Med., 4, 802-807) or neuro-degenerative diseases (Flanders et al., (1998), Prog. Neurobiol., 54, 71-85). The peptides according to the invention containing these motifs can advantageously serve as antagonists by inhibiting the attachment of the TGF-βs to their natural receptors,

- N-glycosylation motifs. The glycosylation of the envelope proteins of retroviruses appears to be directly associated with their functional properties, for example by influencing the number of determinants available in the T cells or by promoting recognition of antigens by the T cells. Glycosylation could play a role in the outbreak or the spread of a pathological condition with an autoimmune component. The glycosylations are necessary for maintaining the conformation of certain epitopes, in particular during the production of a recombinant envelope protein so as to develop a diagnostic reagent and to promote the efficacy of a possible vaccine. Positions 171, 210, 216, 236, 244, 283 and 411. Expected number at random: 3.2

- prenylation sites. Prenylation is an essential mechanism for attachment to the cell membrane and for the targeting of certain proteins. This targeting process could be essential for the production of specific therapeutic agents capable of interfering with the production and regulation of the traffic of cellular complexes calling into play proteins involved in the cell interactions, growth and movement. Positions 188 and 290. Expected number at random: 1.8

- targeting sites in the endoplasmic reticulum. These sites could make it possible to bring about the targeting toward the endoplasmic reticulum in order to carry out the modifications necessary for promoting membrane crossing. Positions 353 and 431. Expected number at random: 0.2

Moreover, the inventors have shown that a number of peptides derived from the env protein of HERV-7q (enverin) have a high affinity/half-life for

the class I HLA alleles. CADD analysis has made it possible to select candidate peptides, for which the best scores are indicated in Table I:

TABLE I

Location	Sequence	HLA molecule	Score	Sequence No.
399	FLGEECCYYV	A-0201	7214	SEQ ID NO: 68
462	LLFGPCIFNL	A-0201	1792	SEQ ID NO: 69
189	CLPLNFRPYV	A-0201	1453	SEQ ID NO: 70
439	GLLSQWMPWI	A-0201	488	SEQ ID NO: 71
263	CLPSGIFFV	A-0201	5103	SEQ ID NO: 72
444	WMPWILPFL	A-0201	897	SEQ ID NO: 73
252	IRWVTPPTQI	B-2705	3000	SEQ ID NO: 74
432	LRNTGPWGLL	B-2705	2000	SEQ ID NO: 75
158	LRTHTRLVSL	B-2705	2000	SEQ ID NO: 76
316	KRVPILPFVI	B-2705	1800	SEQ ID NO: 77
25	CRCMTSSSPY	B-2705	1000	SEQ ID NO: 78
137	TRVHGTSSPY	B-2705	1000	SEQ ID NO: 79
124	AREKHVKEVI	B-2705	600	SEQ ID NO: 80
478	SRIEAVKLQM	B-2705	600	SEQ ID NO: 81
442	SQWMPWILPF	B-2705	500	SEQ ID NO: 82
405	CYYVNQSGI	Kd	2400	SEQ ID NO: 83
346	FYYKLSQEL	Kd	2400	SEQ ID NO: 84
244	TYTTNSQCI	Kd	2400	SEQ ID NO: 85
291	SFLVPPMTI	Kd	1600	SEQ ID NO: 86
406	YYVNQSGIV	Kd	1200	SEQ ID NO: 87
167	LFNTTLTGL	Kd	1152	SEQ ID NO: 88
463	LFGPCIFNL	Kd	960	SEQ ID NO: 89
253	RWVTPPTQI	Kd	480	SEQ ID NO: 90
449	LPFLGPLAAI	B-5102	2200	SEQ ID NO: 91
3	LPYHIFLFTV	B-5102	1210	SEQ ID NO: 92
331	GALGTGIGGI	B-5102	798	SEQ ID NO: 93
321	LPFVIGAGVL	B-5102	550	SEQ ID NO: 94
499	RRPLDRPAS	B-2705	600	SEQ ID NO: 95
194	FRPYVSIPV	B-2705	600	SEQ ID NO: 96
383	RRALDLLTA	B-2705	600	SEQ ID NO: 97
39	WRMQRPGNI	B-2705	600	SEQ ID NO: 98
423	DRIQRRAEEL	B14	1800	SEQ ID NO: 99
158	LRTHTRLVSL	B14	600	SEQ ID NO: 100
359	ERVADSLVTL	B14	540	SEQ ID NO: 101
463	LFGPCIFNLL	Kd	1658	SEQ ID NO: 102
345	QFYYKLSQEL	Kd	1152	SEQ ID NO: 103
443	QWMPWILPFL	Kd	691	SEQ ID NO: 104
405	CYYVNQSGIV	Kd	500	SEQ ID NO: 105
474	NFVSSRIEAV	Kd	480	SEQ ID NO: 106
221	GPLVSNLEI	B-5102	1320	SEQ ID NO: 107
190	LPLNFRPYV	B-5102	726	SEQ ID NO: 108
449	LPFLGPLAAI	B-5101	1144	SEQ ID NO: 109
488	EPKMQSKTKI	B-5101	968	SEQ ID NO: 110
3	LPYHIFLFTV	B-5101	629	SEQ ID NO: 111
125	REKHVKEVI	Kk	1000	SEQ ID NO: 112
312	KPRNKRVPIL	B7	800	SEQ ID NO: 113
378	VVLQNRRAL	Db	792	SEQ ID NO: 114

Location	Sequence	HLA molecule	Score	Sequence No.
377	AVVLQNRRAL	Db	660	SEQ ID NO: 115
321	LPFVIGAGV	B-5101	629	SEQ ID NO: 116
304	DLYSYVISK	A3	540	SEQ ID NO: 117
301	TEQDLYSYVI	Kk	500	SEQ ID NO: 118

This Table I indicates an estimation of the dissociation half-life of a peptide of enverin with an allele of the class I HLA system (the tables of Parker coefficients: J. Immunol, (1994), 152, 163-175). The location indicates the position of the first amino acid of the peptides tested in the enverin sequence. The one-letter code is used for the amino acid sequence. The scores around 500 or greater than 500 were selected. By way of comparison, an analysis was carried out on a concatenation of peptides (polypeptide of 4968 amino acids) reputed to bind the molecules of the class I major histocompatibility complex (Rammensee, Immunogenetics, (1995), 41, 178-228); the ten best scores recorded for nonapeptides and the HLA type A_0201 are respectively 4984, 4047, 2406, 1267, 800, 705, 607, 591, 591 and 577.

It can be seen from this Table I that some molecules of the type I major histocompatibility complex are capable of binding peptides derived from enverin, thus assimilated with peptides of viral or tumor origin, at the level of the endoplasmic reticulum. The complexes formed at the level of the endoplasmic reticulum are then transported to the cell surface, which causes the destruction of the target cell by the cytotoxic T lymphocytes. The peptides identified generally comprise 8 to 10 amino acids. Studies have shown that some alleles of the class I HLA system are thus associated with certain pathologies, in particular with an autoimmune character, such as HLA-B27 with rheumatoid spondylitis or HLA-B51 with Behçet's disease.

A peptide capable of binding a particular class I molecule is consequently capable of functioning as a T cell epitope.

Consequently, the present invention also includes the fragments 399-471 and 244-271 of enverin which advantageously group together several epitopes having high affinity for various haplotypes of the class I HLA system. The use of all or some of these polypeptides is consequently capable of promoting an increase in the T cell repertoire, by allowing better efficacy of the immune response in the context of the various immunotherapeutic, prophylactic or vaccine strategies. These polypeptides may be advantageously delivered for example by the use of viral vectors, viral or synthetic particles, lipopeptides, conventional adjuvants, naked nucleic acids or nucleic acids adsorbed on particles, or liposomes.

For the purposes of the present invention, the peptides may be chemically or biochemically modified; some of the amino acids may be replaced with an analogous amino acid, according to conventional criteria for homologies (A or G; S or T; I, L or V; F, Y or W; N or Q; D or E).

The subject of the present invention is also immunogenic or vaccine compositions for protecting against autoimmune diseases, in particular in at-risk subjects, characterized in that it comprises at least one peptide comprising at least one motif of the CKS type and/or at least one peptide consisting of a motif having affinity with one of the haplotypes of the class I or class II HLA system and a pharmaceutically acceptable vehicle.

According to an advantageous embodiment of said composition, said motif is selected from the group consisting of peptides, as defined in Table I above.

According to another advantageous embodiment of said composition, said peptide has the following sequence:

sequence CKH: LQNRRALDLLTAERGGTcl**FLGEECCYYV** (SEQ ID NO: 120).

It is remarkable to note at the level of position 380 of the enverin protein, the contiguousness

of the motifs of the CKS-17 type (underlined) and of the peptide having the highest score (in bold; see peptide at position 399 in Table I, SEQ ID NO: 68) in the sequence CKH.

The clonal activation of the subgroups of lymphocytes, for example of cytotoxic lymphocytes, by the peptides in Table I and by extension their homologues, is blocked by conventional immunotherapy means such as for example serotherapy and vaccination.

The combination of two sequences or of the sequences analogous to the CKH peptide (SEQ ID NO: 120), is capable of causing a synergistic process in the immune response, which could bring into play additional signaling and activation pathways capable of modulating the lymphocyte activation.

The vaccination relates to the production of antibodies directed against the peptides of Table I, according to the rules of the prior art and according to the methods of release controlled by artificial or cellular implants using a composition as defined above and by using gene therapy means, such as for example expression of nucleic sequences encoding the peptides of Table I. Consequently, the subject of the invention is also immunogenic or vaccine compositions, characterized in that they comprise a vector including at least one nucleic sequence encoding a peptide as defined in Table I, optionally combined with a sequence encoding a motif of the CKS-17 type.

The serotherapy relates to the use of neutralizing antibodies produced from the peptides of Table I and their homologues.

The protein products generated by the endogenous retroviral sequences or produced in parallel may be advantageously characterized by micro-methods of analysis and quantification of peptides and proteins: HPLC/FPLC or equivalent, capillary electrophoresis or equivalent, microsequencing techniques (Edman method or equivalent, mass spectrometry and the like).

The subject of the invention is also antibodies directed against one or more of the peptides described above and their use either for carrying out a method, in particular a differential method, of *in vitro* detection of the presence of such a sequence in an individual, or for the preparation of a composition capable of being used in serotherapy in neuropathological conditions with an autoimmune component.

Said antibodies are advantageously polyclonal or monoclonal antibodies obtained by an immunological reaction from a human, mammalian or avian organism or other species toward the proteins, as defined above.

The subject of the present invention is a method for the differential immunological screening of normal or pathological human endogenous retroviral sequences of the HERV-7q family, characterized in that it comprises bringing a biological sample into contact with an antibody according to the invention, the reading of the result being visualized by an appropriate means, in particular EIA, ELISA, RIA, fluorescence.

By way of illustration, such an *in vitro* diagnostic method according to the invention comprises bringing a biological sample collected from a patient into contact with antibodies according to the invention and detecting with the aid of any appropriate method, in particular with the aid of labeled anti-immunoglobulins, the immunological complexes formed between the proteins produced normally or pathologically and the antibodies.

Monoclonal or polyclonal antibodies, produced from antigens corresponding to synthetic peptides, or recombinant polypeptide or proteins make it possible to monitor the expression of the peptides or proteins produced normally or pathologically. The analysis is preferably carried out by ELISA or equivalent, Western blotting or equivalent, or by immunohistochemistry.

The peptides or proteins, derived from the endogenous retroviral sequences or whose expression is associated with the expression of these endogenous retroviral sequences, are tested for and identified.

The subject of the present invention is also a method for the identification and detection of endogenous retroviral motifs which are abnormally expressed in the context of pathological conditions associated with cancer, or of neuropathological conditions, in particular autoimmune neuropathological conditions, at the forefront of which is multiple sclerosis, characterized in that it comprises the comparative analysis of the sequences extracted from a biological sample and the sequences according to the invention.

The subject of the present invention is also the application of the nucleic sequences or of the protein sequences according to the invention to the diagnosis of, to the prognosis of, to the evaluation of genetic susceptibility to, any induced, congenital or acquired human diseases, in particular those with cancerous, autoimmune and/or neurological components, such as multiple sclerosis, the associated syndromes and the neurodegenerative diseases in which all or part of the nucleic sequences according to the invention and related endogenous or exogenous forms are involved.

The subject of the present invention is also hybrid nucleic sequences, characterized in that they comprise nucleic sequences or motifs according to the invention, combined with sequences or motifs of endogenous origin or of exogenous origin or induced exogenously.

The subject of the present invention is, in addition, a recombinant cloning or expression vector, characterized in that it comprises a nucleic sequence in accordance with the invention.

Therapeutic strategies may be envisaged by using some of the nucleic sequences contained in HERV-7q and the sequences of the same family or deduced

polypeptide structures or by the use of peptides or proteins, or of specific antibodies.

In accordance with the invention, all or part of the endogenous retroviral nucleic sequences of the HERV-7q type may be used for use as a vector or as vector elements for therapeutic use, in particular the LTR sequences and the *gag* region (SEQ ID NO: 2, 21 and 22).

The advantage of such sequences lies in the safety of the vector thus formed, in the possibility of a targeted specific insertion in a well-defined region by a strategy similar to homologous recombination, in cellular targeting, which is optionally transient in the case of a placental expression in women. Another aspect relates to the possibility of combining with the genes of interest the biologically active retroviral motifs (immunomodulatory peptides, as represented in the sequences SEQ ID NO: 68-118, below, fusogenic peptide and the like).

The subject of the present invention is also transgenic animals, characterized in that they comprise all or part of a sequence of the HERV-7q type (SEQ ID NO: 1-22 and 61).

Table II below establishes the correspondences between the sequence numbers as they appear in the sequence listing and the name of the various sequences.

TABLE II

SEQ ID NO:	DESIGNATION
1	Nucleic acid: 7 env
2	Nucleic acid: gag
3	Nucleic acid: HERV-7q
4	Nucleic acid: HE2
5	Nucleic acid: HE3
6	Nucleic acid: HG3
7	Nucleic acid: HE4
8	Nucleic acid: HE5
9	Nucleic acid: HE6
10	Nucleic acid: HG6
11	Nucleic acid: HE7
12	Nucleic acid: HE8
13	Nucleic acid: HG8
14	Nucleic acid: HE9
15	Nucleic acid: HE10
16	Nucleic acid: HE11
17	Nucleic acid: HG11
18	Nucleic acid: HE12
19	Nucleic acid: HG12
20	Nucleic acid: R1
21	Nucleic acid: RIF
22	Nucleic acid + deduced env protein: HERV-7q
23	Fragment of deduced env protein according to SEQ ID NO: 22
24	Fragment of deduced env protein according to SEQ ID NO: 22
25	Fragment of deduced env protein according to SEQ ID NO: 22
26	Protein: enverin
27	Fragment of deduced env protein according to SEQ ID NO: 22
28	Nucleic acid + protein deduced from gag: HERV-7q
29	Fragment of deduced gag protein according to SEQ ID NO: 28
30	Fragment of deduced gag protein according to SEQ ID NO: 28
31	Fragment of deduced gag protein according to SEQ ID NO: 28
32	Fragment of deduced gag protein according to SEQ ID NO: 28
33	Fragment of deduced gag protein according to SEQ ID NO: 28
34	Fragment of deduced gag protein according to SEQ ID NO: 28
35	env protein: reading frame 1
36	gag protein
37	Nucleic acid: G1F (primer)
38	Nucleic acid: G1R (primer)
39	Nucleic acid: G2F (primer)
40	Nucleic acid: G2R (primer)
41	Nucleic acid: G4F (primer)
42	Nucleic acid: G3F (primer)
43	Nucleic acid: G4R (primer)
44	Nucleic acid: G5R (primer)
45	Nucleic acid: E1F (primer)
46	Nucleic acid: E1R (primer)
47	Nucleic acid: E2F (primer)
48	Nucleic acid: E2R (primer)
49	Nucleic acid: E3F (primer)
50	Nucleic acid: E3R (primer)
51	Nucleic acid: E4F (primer)

SEQ ID NO:	DESIGNATION
52	Nucleic acid: E4R (primer)
53	Nucleic acid: E5F (primer)
54	Nucleic acid: E6F (primer)
55	Nucleic acid: E5R (primer)
56	Nucleic acid: ExF (primer)
57	Nucleic acid: ExR (primer)
58	Protein gag
59	Nucleic acid: Sequence A (insertion sequence)
60	Nucleic acid: Sequence B (insertion sequence)
61	Nucleic acid: HE13
62	Nucleic acid: RH7
63	Nucleic acid: RAM75
64	Nucleic acid: RAV73
65	Nucleic acid: RBP3
66	Nucleic acid: HI3
67	Nucleic acid: LTX
68	Peptide Table I
69	Peptide Table I
70	Peptide Table I
71	Peptide Table I
72	Peptide Table I
73	Peptide Table I
74	Peptide Table I
75	Peptide Table I
76	Peptide Table I
77	Peptide Table I
78	Peptide Table I
79	Peptide Table I
80	Peptide Table I
81	Peptide Table I
82	Peptide Table I
83	Peptide Table I
84	Peptide Table I
85	Peptide Table I
86	Peptide Table I
87	Peptide Table I
88	Peptide Table I
89	Peptide Table I
90	Peptide Table I
91	Peptide Table I
92	Peptide Table I
93	Peptide Table I
94	Peptide Table I
95	Peptide Table I
96	Peptide Table I
97	Peptide Table I
98	Peptide Table I
99	Peptide Table I
100	Peptide Table I
101	Peptide Table I
102	Peptide Table I
103	Peptide Table I
104	Peptide Table I
105	Peptide Table I

SEQ ID NO:	DESIGNATION
106	Peptide Table I
107	Peptide Table I
108	Peptide Table I
109	Peptide Table I
110	Peptide Table I
111	Peptide Table I
112	Peptide Table I
113	Peptide Table I
114	Peptide Table I
115	Peptide Table I
116	Peptide Table I
117	Peptide Table I
118	Peptide Table I
119	Nucleic acid: BLIMP-1
120	Peptide: CKH
121	Nucleic acid: F645 (primer)
122	Nucleic acid: PS5D (primer)

In addition to the preceding arrangements, the invention also comprises other arrangements which will emerge from the description which follows, which refers to exemplary embodiments of the method which is the subject of the present invention as well as to the appended drawings, in which:

- Figure 1. Human nucleic sequence HERV-7q, whose analysis and treatment make it possible to characterize a novel endogenous retroviral structure. The repeat nucleic regions of type R1 and R2 and the *gag*, *pol* and *env* domains are underlined. The *gag* and *env* type domains are in italics. The region homologous to a noncoding 3' portion of Rab7 is double underlined.

- Figure 2. Map of the human endogenous retroviral region HERV-7q. The upper part of the figure corresponds to an anonymous region of the human genome situated on the long arm of chromosome 7. The repeat domains (1), *gag* (2), *pol* (3) and *env* (4) of HERV-7q can be identified. The C-terminal *env* region (4.3) is prolonged upstream in the form of a long open reading frame (4.2). The domain 4.1 corresponds to the N-terminal region of the *env* domain.

- Figure 3. Comparison of the repeat nucleic sequences situated at the boundaries of HERV-7q. The 5'

(top) and 3' (bottom) repeat nucleic regions are compared and the identical bases are indicated by two dots.

- Figure 4. Deduced sequence having an open reading frame in the env-type domain of HERV-7q according to the longest open reading frame rule.

- Figure 5. Sequences around the CKS-17 domain identified in various deduced *env* domains of the HERV-7q family and comparison with reference CKS-17 motifs.

1) HE2 - 2) HERV-7q - 3) GenBank accession No.: M85205 - 4) HE7 - 5) HE9 - 6) CKS-17; the peptide motif endowed with immunomodulatory properties is underlined - 7) gp20 of retrovirus type D (SRV-Pc).

- Figure 6. Possible deduced sequence of the *gag*-type domain identified in HERV-7q established according to the longest open reading frame rule. X and / correspond to a non-sense codon and to a reading frame shift, respectively. The underlined sequence corresponds to the beginning of the *pol* domain.

- Figure 7. Comparison of the nucleic regions covering the *gag* region of HERV-7q (top) and HERV-TcR (bottom) and their flanking regions. The identical bases are specified by two dots.

- Figure 8. Example of nucleic alignments of the *env*-type domain of HERV-7q with similar *env*-type domains present in human endogenous retroviral sequences of the same family. The non-sense codons are underlined: 1) HERV-7q - 2) HE2 03) HE3 - 04) HE4.

- Figure 9. Nucleic alignments between the *gag* domain of HERV-7q and the corresponding domains belonging to the same family. Comparison with fragments of *gag* domains isolated from infectious retroviral agents. Sequences of infectious retroviral origin: EMBL database accession No.: 1) A60168 - 2) A60201 - 3) A60200 - 4) A60171. Human endogenous retroviral sequences: 5) HERV-7q - 6) HG11 - 7) HG3. The figures indicated in the endogenous sequences correspond to the number of nucleotides inserted in order to optimize the

alignment with the *gag*-type sequences identified in retroviruses of infectious origin.

- Figure 10. Alignment of a deduced *gag* protein motif (top) belonging to an infectious retrovirus (EMBL accession No.: A60200) with the deduced *gag* protein motif (bottom) identified in HERV-7q. The non-sense codons are in bold and underlined. The identical amino acids are specified by 2 dashes. One dash indicates a deletion or a homologous amino acid.

- Figure 11. Alignment of an *env* motif (top) belonging to an infectious retrovirus (EMBL accession No.: A60170) with the *env* motif (bottom) identified in HERV-7q. The homologous nucleotides are specified by two dots and the deletions by a dash.

- Figure 12. Comparison between the *env* domain of HERV-7q (top) and the *env* domain of HERV-9 (bottom). The 66% homology is limited to the 3' region of the *env* domain of HERV-7q and HERV-9, respectively between nucleotides 8976 nt and 9500 nt of HERV-7q and nucleotides 2898 nt and 3465 nt of HERV-9 (GenBank accession No.: X57147). Numerous insertions/deletions are also observed.

- Figure 13. Homology between a portion of the sequence of the transcript encoding RH7 (top, SEQ ID NO: 62) and an RGH2 motif (bottom - GenBank accession No.: D11018).

-Figure 14. Identification of the sequence of the transcript encoding RAM75 (SEQ ID NO: 63), corresponding to the gene for an ATPase of PEX1 type. The coding exons are underlined. The initiation and non-sense codons as well as the putative poly-adenylation sites are in bold and underlined. The region in italics corresponds to the beginning of the endogenous retroviral sequence RH7.

- Figure 15. Sequence of the transcript encoding RAV73 (SEQ ID NO: 64), situated at 0.7 kb downstream of HERV-7q; the nucleic sequences capable of encoding one or more polypeptides are underlined.

- Figure 16. Comparison between the 3' LTR sequence (top) of HERV-7q and the intron sequence LTX (SEQ ID NO: 67), situated in the FMR2 gene, associated with fragile X (bottom).

- Figure 17. Detection of modifications on the nucleotide sequence (ID NO: 3), in patients suffering from MS. The modified bases, in at least one patient, are underlined. The primers used are in italics (sequences SEQ ID NO: 121 and 122). The initiation ATG and the non-sense codon are in bold.

- Figure 18. The *env* coding portion of the HERV-7q sequence (sequence ID NO: 3), with 3 reading frames.

- Figures 19, 20, 21. Separate presentation of the env protein according to the 3 reading frames.

- Figure 22. Nucleic sequence containing the retroviral sequence RH7 situated in 5' of the HERV-7q sequence. The sequence in italics corresponds to the beginning of the HERV-7q sequence. The RH7 sequence is underlined. Two putative polyadenylation sites are in bold.

- Figure 23. Sequence of the transcript encoding RBP3 containing nucleotide motifs identified in the nucleic sequence encoding the Blimp-1 gene.

- Figure 24. Sequence of the transcript encoding APS.

- Figure 25. Sequence of the transcript encoding Blimp-1; the coding portion is underlined; the initiation and termination codons are in bold.

- Figure 26. Sequence of the transcript encoding FMR2. The coding portion is underlined. The initiation and non-sense codons are in bold.

It should be clearly understood, however, that these examples are given solely by way of illustration of the subject of the invention and do not in any manner constitute a limitation thereto.

EXAMPLE 1: Detection, by gene amplification, of a nucleic sequence belonging to a domain of the *gag* or *env* type according to the invention, in a genomic DNA sample of human or mammalian origin

The gene amplification is carried out using genomic DNA isolated from blood. An anticoagulant treatment is carried out with 1 ml of a citrate solution (per liter: 4.8 g of citric acid, 13.2 g of sodium citrate, 14.7 g of glucose) per 6 ml of fresh blood. After centrifugation of 20 ml of blood for 15 min at 130 000 g, the supernatant is removed and the fraction enriched with white blood cells is transferred into a new tube and then recentrifuged under the same conditions as above. The fraction enriched with white blood cells is resuspended in an extraction buffer (10 nM Tris-HCl, 0.1 M EDTA, 20 µg/ml of pancreatic RNAse treated so as to eliminate the DNAses, 0.5% SDS, pH 8.0), and then incubated for 1 hour at 37°C. Proteinase K is added at a final concentration of 100 µg/ml. The suspension of lyzed cells is incubated at 50°C for 3 hours, with occasional stirring, and then treated with an equal volume of phenol equilibrated with 0.5 M Tris-HCl, pH 8.0. The emulsion formed is placed on a wheel for one hour and then centrifuged at 5 000 g for 15 min at room temperature. The aqueous solution is treated and deproteinized by a triple phenol extraction in order to obtain a level of purification corresponding to an absorbance A260/A280 final ratio greater than 1.75. The aqueous fraction is precipitated with 0.2 vol. of 10 M sodium acetate and 2 vol. of ethanol. The DNA is then either collected with the tip of a bent Pasteur pipette, or centrifuged at 5 000 g for 5 min at room temperature. The DNA or the DNA pellet is washed twice with 70% ethanol and then taken up in 1 ml of TE, pH 8.0 so as to be eluted, with gentle stirring, for 12 to 24 hours.

Oligonucleotides specific for the endogenous sequences described according to the invention are chosen in order to amplify the *gag* or *env* region of the

endogenous retroviral regions described according to the invention. The genomic DNA studied is obtained from patients having pathological conditions such as multiple sclerosis and from individuals reputed to be healthy.

The thermostable DNA polymerases used were chosen for their high accuracy during the amplification process, such as Vent DNA polymerase (Biolabs) and the like, and are used according to the conditions recommended by the supplier.

The amplification strategy uses, depending on the case, a simple PCR, or a nested or seminested PCR.

Oligonucleotides used to amplify the *gag* region:

- primer G1F, sense, located in the region upstream of the *gag* domain of *HERV*-7q (SEQ ID NO: 37),
- primer G1R, antisense, located in the 3' terminal region of the *gag* domain (SEQ ID NO: 38).

The fragment of 1505 nt amplified by the pair G1F-G1R; 1505 nt is used to generate the probes capable of hybridizing the various PCR amplification products.

- primer G2F, sense nested (SEQ ID NO: 39),
- primer G2R, antisense nested (SEQ ID NO: 40),
- primer G4F, sense nested (SEQ ID NO: 41),
- primer G3F, sense nested (SEQ ID NO: 42),
- primer G4R, antisense nested (SEQ ID NO: 43),
- primer G5R, antisense nested (SEQ ID NO: 44).

Oligonucleotides used to amplify the *env* region of HERV-7q:

- primer E1F, sense (SEQ ID NO: 45),
- primer E1R, antisense (SEQ ID NO: 46).

The fragment of 2529 nt amplified by the pair of primers E1F-E1R is used to generate the probes capable of hybridizing the various PCR amplification products.

- primer E2F, sense (SEQ ID NO: 47),
- primer E2R, antisense (SEQ ID NO: 48),
- primer E3F, sense (SEQ ID NO: 49),
- primer E3R, antisense (SEQ ID NO: 50),

- primer E4F, sense (SEQ ID NO: 51),
- primer E4R, antisense (SEQ ID NO: 52),
- primer E5F, sense (SEQ ID NO: 53),
- primer E6F, sense (SEQ ID NO: 54),
- primer E5R (SEQ ID NO: 55),
- primer ExF (SEQ ID NO: 56),
- primer ExR (SEQ ID NO: 57).

The PCR is carried out using 50 to 200 ng of genomic DNA. The PCR conditions are those recommended by the supplier. The amplification cycle conditions are carried out in 50 μl: denaturation of 94°C for 1 min, hybridization of 70°C for 1 min, and extension at 72°C for 1 to 2 min, depending on the amplified fragments. After 35 cycles, a terminal reaction is carried out at 72°C for 10 min. Automated sequencing of the amplified samples is carried out with the aid of an Applied Biosystems type ABI 377 sequencer or another comparable model, according to the protocols provided by the manufacturer.

In the case of a nested or seminested PCR, the same experimental conditions are used, the only difference being that the genomic DNA sequence is replaced with 5 to 10 μl of the amplification product derived from the first PCR.

Two independent amplifications are carried out using the same sample. A control reaction is carried out by replacing the DNA sample with water in order to detect possible contaminants.

EXAMPLE 2: Detection, by gene amplification, of a nucleic sequence according to the invention in a biological sample of genomic DNA collected from patients having an existing candidate pathological condition or suspected of having this pathological condition

The amplification protocol is the same as in Example 1, apart from the origin of the sample which is obtained from patients having a candidate pathological condition. A genomic DNA sample reputed to be normal is

systematically integrated into the set of amplified pathological samples and then analyzed.

The PCR products are separated on a 1.5% agarose gel and then transferred in the presence of 0.4 N sodium hydroxide on a charged nylon membrane. Hybridization is carried out with a specific probe corresponding to the PCR fragments amplified either with the pair G1F-G1R or the pair E1F-E1R. The probe is labeled by incorporating dUTP-digoxygenin according to the supplier's protocol (Boehringer Mannheim). The hybridization is carried out in a hybridization buffer (5XSSC, 50% formamide, 0.1% lauroylsarcosine, 0.02% SDS, 2% blocking reagent Boehringer) overnight at 42°C. The Southern is washed for twice 5 min at room temperature in a 2XSSC solution containing 0.1% SDS. Next, a high stringency wash is carried out twice for 15 min at 55°C in a 0.1XSSC solution containing 0.1% SDS. The hybridization is visualized according to the supplier's protocol (Boehringer Mannheim), in the presence of a chemiluminescent substrate for alkaline phosphatase, of the CSPD or CDP-STAR type. The filter is visualized after a 15 min exposure at 60°C.

SSCP (*single strand conformation polymorphism*) analysis makes it possible to detect discrete modifications of the sequence of the fragments amplified by PCR. The PCR is carried out in the presence of dCTP labeled with ^{32}P. The sample to be analyzed is denatured at 95°C for 10 min in the presence of loading buffer, and then immediately loaded onto a 10% polyacrylamide gel containing 7.5% glycerol. The migration is carried out at 4°C at 8-10 W. The gel is dried and then autoradiographed.

The PCR fragments likely to exhibit an alteration of their nucleotide sequence are sequenced according to Example 1.

Hybridization with the aid of a specific oligonucleotide (17 mers to 20 mers) corresponding to the modified nucleotide region makes it possible to identify the samples having an identical modification

(ASO method). Briefly, the southern is hybridized with an oligonucleotide which is distally labeled either with ^{32}P, or in the presence of digoxygenin (according to the Boehringer Mannheim protocol) and then washed under stringent conditions at 65°C in a 6XSSC solution containing 0.05% sodium pyrophosphate.

For example, an automated nucleotide sequencing was carried out on six PCR fragments obtained from 5 patients suffering from MS and a control reputed to be normal, and which were amplified using the primers F645: CTTCAAACAACAACCAGGAGG (SEQ ID NO: 121) (situated 26 nucleotides upstream of the initiation methionine of enverin) and PS5D: TTGGGGAGGTTGGCCGACGA (SEQ ID NO: 122) (situated 6 nucleotides downstream of the non-sense codon of enverin). Modifications of the sequence of enverin were observed on the DNA from some patients (Figure 17).

EXAMPLE 3: Detection of a protein according to the invention in a biological sample

- Preparation of a purified protein fraction of cerebrospinal fluid from patients suffering from MS

After a treatment at 56°C for 30 min and removal of the immunoglobulins on a G HiTrap protein column (Pharmacia), the equivalent of 10 ml of CSF is deposited on a DEAE Sepharose CL-6B column (Pharmacia). The elution is carried out in 20 mM Tris-HCl, pH 8.8, and a gradient from 0 to 0.4 M NaCl, and then the fraction is dialyzed twice against a phosphate-NaCl buffer (PBS). After concentration on Ultrafree-MC (Millipore), the fraction is deposited on a Superose 12 column (FPLC Pharmacia) and eluted in the presence of PBS. After separation by polyacrylamide-SDS gel electrophoresis and electrotransfer onto an Immobilon-P membrane (Millipore), the protein bands are subjected to controlled trypsin hydrolysis.

- Analysis of the protein fraction by mass spectrometry

The peptides digested in the presence of trypsin are analyzed by the MALDI-TOF method, which

allows the analysis of peptides present in a mixture (COTTRELL J.S., Pept. Res., 1997, 7, 115-124). The peptides characterized according to their mass are compared with the proteins and with the associated proteins according to the invention.

EXAMPLE 4: Detection of specific antibodies to the *env* domain of HERV-7q

The identification of a long open reading frame in the *env* sequence of HERV-7q made it possible to determine a deduced protein sequence SEQ ID NO: 22 and 35 and Figures 18-20 of a region of the said gene.

The protein sequences deduced from the sequences ID NO: 22, 35 and Figures 18-20 are positioned as follows with respect to Figure 1 or the sequence ID NO: 3:

SEQ ID NO: 22 (reading frame 1) and Figure 19: beginning of the coding sequence: position 7874, end of the coding sequence 1st nonsense codon (position 9493)

SEQ ID NO: 35: beginning of the coding sequence: position 7874, end of the coding sequence 1st nonsense codon (position 9493) (reading frame 1)

Figure 19: beginning of the coding sequence: position 6970, end of the coding sequence 1st nonsense codon (position 9493) (reading frame 1)

Figure 20: beginning of the coding sequence: position 6971, the end of the reading frame is shifted depending on the case by 1, 2 or 3 codons

Figure 21: beginning of the coding sequence: position 6972, the end of the reading frame is shifted depending on the case by 1, 2 or 3 codons

Various peptides corresponding to all or part of SEQ ID NO: 22 (see SEQ ID NO: 23-27 and 35) were synthesized by genetic engineering in order to test their antigenic specificity toward sera or tissues from patients suffering from MS, for example. Briefly, all or part of the env region of HERV-7q is subcloned into the vectors pQE30, 31 and 32. The vectors pQE30, 31 and 32 contain, in 5' of the multiple cloning site, the consensus sequences for transcription (the strong T5

bacteriophage promoter, 2 operators of the lactose operon) and translation (one synthetic ribosome binding site). Likewise, pQE30, 31 and 32 possess, in 3', the phage 1 transcription terminator as well as a Stop codon for translation. The expression of the protein is carried out after transformation in *E. coli* M15. The plasmid pQE30, 31 and 32 possess, upstream of the multiple cloning site, the coding sequence for a succession of 6 histidines having affinity for nickel ions. This stretch allows the purification of the expressed chimeric protein by adsorption on a resin consisting of a chelating ligand, nitrotriacetic acid (NTA), charged with 4 nickel ions (NI-NTA resin, Qiagen).

The transformation is carried out by electroporation or treatment with calcium chloride. For example, an *E. coli* M15 colony is incubated in 100 ml of LB medium containing 250 µg of kanamycin, with stirring at 37°C until an OD^{600} of 0.5 is obtained. After centrifugation for 5 minutes at 2000 g at 4°C, the bacterial pellet is taken up in 30 ml of TFB1 solution (100 mM rubidium chloride, 50 mM manganese chloride, 30 mM potassium acetate, 10 mM $CaCl_2$, 15% glycerol, pH 5.8), at 4°C for 90 minutes. After a centrifugation of 5 minutes at 2000 g at 4°C, the bacterial pellet is taken up in 4 ml of TFB2 solution (10 mM rubidium chloride, 10 mM MOPS, 75 mM $CaCl_2$, 15% glycerol, pH 8). The cells may be kept at -70°C in aliquots of 500 ml. 20 µl of the ligation and 125 µl of competent cells are mixed and placed on ice for 20 minutes. After a heat shock of 42°C for 90 seconds, the cells are stirred for 90 minutes at 37°C in 500 ml of Psi-broth medium (LB medium supplemented with 4 mM $MgSO_4$, 10 mM potassium chloride). The transformed cells are plated on LB-agar dishes supplemented with 25 µg/ml of kanamycin and 100 µg/ml of ampicillin, and the dishes are incubated overnight at 37°C.

The potentially recombinant clones are subcultured in an orderly manner on a nylon filter

deposited on an LB-agar dish supplemented with 25 μg/ml of kanamycin and 100 μg/ml of ampicillin. After one night at 37°C, the recombinant clones are located by hybridization of the plasmid DNA with the nucleotide probe amplified by PCR with the pair of primers according to SEQ ID NO: 45 and SEQ ID NO: 46.

An independent colony containing the insert is inoculated at 20 ml of LB medium supplemented with 25 μg/ml of kanamycin and 100 μg/ml of ampicillin. After one night at 37°C, with stirring, 500 ml of the same medium are incubated at 1/50 with this preculture until an OD^{600} of 0.8 is obtained, and then 1 to 2 mM final of IPTG is added. After 5 hours, the cells are centrifuged for 20 minutes at 4 000 g.

A portion of the cellular pellet is taken up in 5 ml of sonification buffer (50 mM of sodium phosphate, pH 7.8, 300 mM NaCl) and then placed on ice. After rapid sonification, the cells are centrifuged for 20 minutes at 10 000 g. A portion of the cellular pellet is taken up in 10 ml of a 30 mM Tris/HCl-20% sucrose solution pH 8. The cells are incubated for 5 to 10 minutes, with stirring, after addition of 1 mM EDTA. After a centrifugation of 10 minutes at 8 000 g at 4°C, the pellet is taken up in 10 ml of 5 mM ice cold $MgSO_4$. After 10 minutes on the ice, with stirring, the cells are centrifuged for 10 minutes at 8 000 g at 4°C.

The pellet is taken up in 5 ml/g in buffer A (6 M GuHCl (guanidine hydrochloride), 0.1 M sodium phosphate, 0.01 M Tris/HCl, pH 8), 1 hour at room temperature. The lysate is centrifuged for 15 minutes at 10 000 g at 4°C, and the supernatant is supplemented with 8 ml of Ni-NTA resin, pre-equilibrated in buffer A. After 45 minutes at room temperature, the resin is poured into a column, washed with 10 times the column volume with buffer A and then with 5 times the column volume with buffer B (8 M urea, 0.1 M sodium phosphate, 0.01 M Tris/HCl, pH 8). The column is washed with buffer C (8 M urea, 0.1 M sodium phosphate, 0.01 M Tris/HCl, pH 6.3) until A280 is less than 0.01. The

recombinant protein is eluted with 10 to 20 ml of buffer D (8 M urea, 0.1 M sodium phosphate, 0.01 M Tris/HCl, pH 5.9) and then with 10 to 20 ml of buffer E (8 M urea, 0.1 M sodium phosphate, 0.01 M Tris/HCl, pH 4.5), and then with 20 ml of buffer F (6 M HCl, 0.2 M acetic acid). After SDS-PAGE analysis, the purified fraction(s) containing the chimeric protein allowed the production of antibodies in rabbits. The antibodies obtained are tested by Western blotting after visualization with a secondary antibody coupled to alkaline phosphatase.

Antibodies are obtained in the same manner, using peptides synthesized chemically according to the Merrifield technique (G. Barany and B. Merrifield, 1980, in *The peptides*, 2, 1-284, E. Gross and J. Meienhofer, Academic Press, New York).

The specific antibodies obtained are used for detection of the serum or tissue expression of all or part of the endogenous retroviral sequences according to the invention, in normal and pathological cases.

The proteins of serum or tissue origin are separated on acrylamide-SDS gel and then transferred onto a nitrocellulose filter with the aid of a Novablot 2117-2250 apparatus (LKB). The transfer is carried out on a Hybond C-extra sheet (Amersham) using a 100 mM CAPS buffer pH 11, methanol, water (V/V/V: 1/1/8) containing 1 mM $CaCl_2$. After a transfer of 1 hour at 0.8 mA/cm^2, the sheet is saturated for 1 hour at room temperature in PBS-0.5% gelatin. The sheet is brought into contact with the specific antibody at the concentration of 1/1 000 in PBS-0.25% gelatin. After 2 hours, the filter is washed 3 times 15 minutes in PBS-0.1% Tween-20, and then the filter is incubated for 30 minutes in the presence of a secondary antibody coupled to alkaline phosphatase (Promega), diluted 1/7 500 in PBS-0.25% gelatin. After three washes in PBS-0.1% Tween-20, the filter is equilibrated in a buffer (100 mM Tris-HCl, pH 9.5, 100 mM NaCl, 5 mM $MgCl_2$). The visualization is carried out in the presence

of 45 µl of NBT at 75 mg/ml and 35 µl of BCIP at 50 mg/ml, per 10 ml of alkaline phosphatase buffer.

The chimeric proteins obtained by genetic engineering are also used for tests of biological activity, such as for example the test for biological activity of the CKS-17-type peptide identified in the *env* domain of HERV-7q (Figure 5).

EXAMPLE 5: Production of ribonucleic probes encoding the *env* sequences of HERV-7q

The PCR fragments obtained are subcloned into the plasmid PGEM 4Z (Promega) which possesses on either side of its multiple cloning site, promoter sequences for the SP6 and T7 RNA polymerases.

The method of competence used is electroporation. The plasmid and the PCR fragment are hybridized in a ratio of 50 ng of vector (SmaI cleavage) to 100 ng of PCR fragment (made blunt ended by treatment with the Klenow fragment of DNA polymerase). The incubation takes place overnight at 22°C in ligation buffer (66 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 1 mM dithioerythritol, 1 mM ATP) in the presence of 1 u of T4 DNA ligase and is then stopped by denaturation for 10 minutes at 65°C. In parallel, the *E. coli* JM 105 strain is inoculated overnight at 37°C in LB medium. This preculture is diluted 1/500 and placed at 37°C until an OD^{600} equal to 1 is obtained. For the remainder of the procedure, the cells will always be stored at cold temperature. After centrifugation for 5 minutes at 3 500 g at 4°C, the cellular pellet is resuspended in 1/4 vol. of ultrapure ice-cold water. This step is repeated 5 to 6 times. The pellet is then resuspended in 1/4 000 vol. of water; 10% of sterile glycerol is added, allowing preservation of the electrocompetent cells, in aliquots of 10 µl at 20°C. 1 µl of the ligation is added to 50 µl of electrocompetent cells; the mixture is subjected to an electrical discharge of 12.5 kV/cm, applied for 5.8 ms. The cells are rapidly resuspended in the SOC medium, incubated for 1 hour at 37°C and

then plated in the presence of 2% X-Gal in dimethylformamide, and 10 mM IPTG, on an LB-agar dish supplemented with ampicillin (100 µg/ml). After one night at 37°C, the potentially recombinant white clones are subcultured in an orderly manner on an LB/ampicillin dish and in parallel on a nylon filter deposited on an LB/ampicillin dish. These two dishes are incubated overnight at 37°C. The recombinant clones are then located by hybridization with a nucleic probe amplified by PCR with the pair or primers according to SEQ ID NO: 45 and SEQ ID NO: 46 and labeled with digoxygenin.

The recombinant clones are cultured in 50 ml of LB/ampicillin medium (100 µg/ml), with stirring, overnight at 37°C. After centrifugation at 3 500 g for 15 minutes at 4°C, the bacterial pellet is taken up in 4 ml of P1 buffer (50 mM Tris-HCl, 10 mM EDTA, 400 µg/ml RNase A, pH 8) and 4 ml of P2 buffer (200 mM NaOH, 1% SDS). The medium is incubated at room temperature for 5 minutes. After addition of 4 ml of P3 buffer (2.55 M potassium acetate, pH 4.8), the mixture is centrifuged at 12 000 g for 30 minutes at 4°C. This supernatant is applied to a Qiagen type 100 column, pre-equilibrated with 2 ml of QBT buffer (750 mM NaCl, 50 mM MOPS, 15% ethanol, pH 7), the column is washed with twice 4 ml of QC buffer (1 M NaCl, 50 mM MOPS, 15% ethanol, pH 7) and the DNA is eluted with 2 ml of QF buffer (1.2 M NaCl, 50 mM MPOS, 15% ethanol, pH 8). The DNA is precipitated with 0.8 vol. of isopropanol and centrifuged at 12 000 g at 4°C for 30 minutes. The pellet is washed with 70% ice-cold ethanol and then the plasmid DNA is taken up in twice 150 µl of TE buffer.

The ribonucleic probes are used as specific probes, in particular for the detection of the transcripts expressed by the endogenous retroviral sequences according to the invention.

EXAMPLE 6: Construction of a transgenic mouse containing all or part of the gene for enverin

A transgenic mouse containing all or part of the HERV-7q sequence (SEQ ID NO: 3) is constructed so as to identify the sequences responsible for the tissue specificity, and to evaluate the role of all or part of the endogenous retroviral motifs of the HERV-7q type, in particular all or part of the peptide motifs of enverin. The microinjection technique used refers to the conventional technique (Hogan et al., (1994), Manipulating the mouse embryo, Cold Spring Harbor, Cold Spring Harbor Laboratory Press) or to its equivalents. Forms identical to the normal human molecule of motifs of the HERV-7q type, including enverin, or forms which are mutated, deleted, having insertions, or truncated are tested in order to determine the motifs which are critical both from the normal and pathological point of view, and more particularly during fetal development and during tumor processes.

Bibliographic references:

- Benit L. et al., 1997. Cloning of a new murine endogenous retrovirus MuERV-L, with strong similarity of the human HERV-L element and with a *gag* coding sequence closely related to the Fv1 restriction gene. J. Virol. 71, 5652-5657.

- Coffin J.M. 1985. Endogenous retrovirus, In: "RNA tumor viruses" (Weiss R.A., Varmus H.E., Teich N.M., and Coffin J.M. eds), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

- Conrad B., Weissmahr R.N., Boni J., Arcari R., Schupbach J., and Mach B. 1997. A human endogenous retroviral superantigen as candidate autoimmunogene in type 1 diabetes. Cell 90, 303-313.

- Covey S.N. 1986. Amino acid sequence homology in *gag* region of reverse transcribing elements and the coat protein gene of cauliflower mosaic virus, Nucleic Acids Res. 14, 623-633.

- Hertig C., Coupar B.E., Gould A.R., and Boyle D.B. 1997. Field and vaccine strains of fowlpox virus carry

integrated sequences from the avian retrovirus, reticuloendotheliosis virus. Virology 235, 367-376.

- Hohenadl C., Leib-Mösch C., Hehlemann R., and Erfle Y. 1996. Biological significance of human endogenous retroviral sequences. J. Acqui. Imm. Def. Synd. Hum. Retrovir. 13, S268-S273.

- Kulkoski J.K., Jones S., Katz R.A., Mack J.P.G., and Skalka A.M. 1992. Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. Mol. Cell. Biol. 12, 2331-2338.

- La Mantia G. et al., N.A.R., 1991, 19, 7, 1513-1520

- Patience C., Wilkinson D.A., and Weiss R.A. 1997. Our retroviral heritage. Trends Genet. 13, 116-120.

- Pearson W.R. 1994. Using the FASTA program to search protein and DNA sequence databases. Methods Mol. Biol. 24, 307-331.

- Perron H., Garson J.A., Bedin F., Beseme F., Paranhos-Baccala G., Komurian-Pradel F., Mallet F., Tuke P.W., Voisset C., Blond J.L., Lalande B., Seigneurin J.M., Mandrand B. and the Collaborative Research Group on Multiple Scelerosis. 1997. Molecular identification of a novel retrovirus repeatedly isolated from patients with multiple sclerosis. Proc. Natl. Acad. Sci. USA 94, 7583-7588.

- Tönjes R.R. et al., J. AIDS and Hum. Retrovirol. 1996, 13. S261-S267.

- Vitelli R., Chiarillo M., Lattero D., Bruni C.B., and Bucci C., 1996. Molecular cloning and expression analysis of the human Rab7 GTP-ase complementary deoxyribonucleic acid. Biochem. Biophys. Res. Commun. 229, 887-890.

- Weber L.T., Miller M., Jaskolski M., Leis J., Skalka M., and Wlodawer A., 1989. Molecular modeling of the HIV-1 protease and its substrate binding site. Science 243, 928-931.

- Wilkinson D., Mager D.L., and Leong J.A.C. 1994. Endogenous human retroviruses. In: "The Retroviridae"

(Levy J.A. ed). Plenum Press New York., Vol. 3, 465-535.

- Xiong Y., and Eickbush, T. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 9, 3353-3362.

As is evident from the above, the invention is not at all limited to its embodiments, implementations and applications which have just been described more explicitly; it embraces on the contrary all the variants which may occur to a specialist in this field, without departing from the framework or scope of the present invention.